In the name of Allah, the Most Gracious, the Most Merciful

LA FACULTÉ

SEEK KNOWLEDGE

Copyright disclaimer

"La faculté" is a website that collects copyrights-free medical documents for non-lucratif use.
Some articles are subject to the author's copyrights.
Our team does not own copyrights for some content we publish.
"La faculté" team tries to get a permission to publish any content; however , we are not able to contact all the authors.
If you are the author or copyrights owner of any kind of content on our website, please contact us on: facadm16@gmail.com
All users must know that "La faculté" team cannot be responsible anyway of any violation of the authors' copyrights.
Any lucrative use without permission of the copyrights' owner may expose the user to legal follow-up.

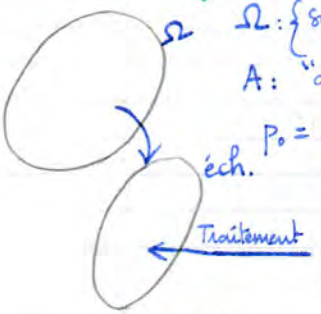
Chapitre 2: Tests d'hypothèse ou tests statistiques

A/ Généralités:

I/ Utilité des tests statistiques en santé:

1/ Évaluation de l'efficacité d'un traitement

ex:



Ω: {souris cancéreuses}

A: "avoir le cancer"

$p_0 = P(A) = 20\%$ dans la pop.

$p = P(A)$

i/ $p_0 = 20\%$; $p = 20\%$; $p = p_0$.

Il n'y a pas de différence entre p et p_0

ii/ $p_0 = 20\%$; $p = 21\%$

Il n'y a pas d'évidence de différence entre p et p_0.

La différence peut être dûe au hasard (fluctuation de l'éch.).

iii/ $p_0 = 20\%$; $p = 1\%$

Il y a une évidence de différence entre les 2 proportions.

iv/ $p_0 = 20\%$; $p = 15\%$? → Test.

On déclarera le traitement actif si, le taux de cancéreux après traitement p s'écarte "nettement" du taux de cancéreux avant le traitement p_0. C'est le sens que l'on donne à ce "nettement" qui est le fondement du principe des tests statistiques.

On suppose 2 hypothèses:

H_0 : $p = p_0$ contre H_1 : $p \neq p_0$ ($p > p_0$ ou $p < p_0$)

H_0 : hypothèse nulle.

H_1 : hypothèse alternative.

2/ Évaluation des facteurs de risque:

Utilisateurs de tél mobile 66/469 ≃ 14%	Tumeur cérébrale n=469
Utilisateurs de tél mob 76/422 ≃ 18%	Pas de tumeur n=422

Comment interpréter la proportion un peu plus élevée d'utilisateurs de portables chez les non-malades?

Réelle association entre l'utilisation du portable et tumeur cérébrale ?

3/ Finalité des tests statistiques:

- Lire et interpréter une étude.
- Tester l'efficacité et la sécurité d'un traitement.

II/ Principe général des tests:

1/ Mise en oeuvre des tests:

a/ Choisir clairement l'hypothèse nulle H_0. En général (toujours), H_0 engage une réalité.

H_1 est le complémentaire de H_0

b/ Les tests stat. sont basés sur des lois de distribution théoriques (Normale, Student, Khi-deux).

(étape calculatoire).

On construit une v.a. à partir d'un éch. choisi. Donc, cette v.a. est fonction des observations.

ex: $Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow N(0,1)$.

Ces lois nécessitent les conditions d'application.

$T = \frac{\bar{X} - m}{\frac{S}{\sqrt{n-1}}} \rightsquigarrow St_{(n-1)}$

Cette variable aléatoire est appelée Statistique de test.

c/ Conclusion: définir une règle de décision qui est soit non rejet de H_0 ou rejet de H_0, avec un risque α.

2/ Différents types d'erreurs:

a/ Erreur de 1ère espèce ou de type I:

$\alpha = P$(rejeter H_0 alors que dans la réalité, H_0 est vraie).

$\alpha = P$(conclure H_1 alors que H_0 est vraie).

α est fixé au préalable ($\alpha = 5\%$).

b/ Erreur de 2nde espèce ou de type II:

$\beta = P$(rejeter H_1 alors que H_1 est vraie)

$\beta = P$(conclure H_0 alors que H_1 est vraie)

N.B.:

H_0 et H_1 ne sont pas symétriques.

α et β n'ont pas le même poids.

Il est plus grave de commettre l'erreur α que l'erreur β.

ex: Dans un essai thérapeutique, où le but est d'établir l'efficacité d'un traitement, le risque d'erreur le plus grave est de déclarer:

$\alpha = P$(le trait. est efficace (H_1) alors qu'il ne l'est pas (H_0)).

$\beta = P$(le trait. est non efficace alors qu'il l'est).

3/ Puissance:

Elle est définie par:

$\pi = 1 - \beta$.

$\pi = 1 - P$(rejeter H_1 alors que H_1 est vraie)

$\pi = P$(ne pas rejeter H_1 alors que H_1 est vraie)

4/ Degré de signification / pvaleur / p-value:

$\alpha = 5\%$: conclusion → rejeter H_0

$\alpha = 4\%$: conclusion → rejeter H_0

$\alpha = 3\%$: conclusion → rejeter H_0

⋮ ⋮

$\alpha = \alpha_0$: conclusion → ne pas rejeter H_0

α_0 est la plus petite erreur (erreur minimale) avec laquelle H_0 est rejetée ou l'erreur maximale avec laquelle H_0 est non rejetée.

N.B.: La p.valeur mesure la force avec laquelle H_0 est rejetée, plus elle petite et plus est confortable le rejet de H_0.

α est fixé.

α_0 est calculé à partir de l'éch. choisi.

Soit Z la stat de test, et soit z_0 la valeur observée de Z sur l'éch. choisi.

$$\alpha_0 = \begin{cases} 2F(z_0) & \text{si } F(z_0) < \frac{1}{2} \\ 2(1 - F(z_0)) & \text{si } F(z_0) \geq \frac{1}{2} \end{cases}$$

où F la fonction de répartion de Z

$$F(z_0) = P(Z \leq z_0).$$

α_0 ne se calcule que pour H_0 rejetée.

B/ Tests de conformité:

On compare une pop. avec un éch.

I/ Test pour la moyenne:

Le but est de comparer la moyenne m connue de la pop. à la moyenne $\bar{x}$ de l'éch.

1ère étape:

H_0 : $m = \bar{x} \Leftrightarrow |m - \bar{x}| \simeq 0$.

H_1 : $m \neq \bar{x}$

H_0: il n'y a pas de différence significative entre les moyennes à comparer l'éch. choisi est représentatif de la pop.

2ème étape:

1ère méthode:

Basée sur les IC ou les IP (intervalle de Pari).

$IC_\alpha(m) = [\bar{x} - h ; \bar{x} + h]$

P $\bar{x}$ m m

IC { m ; $\bar{x}-h$ $\bar{x}$ $\bar{x}+h$

IP { $\bar{x}$; $m-h$ m $m+h$

a/ Déf:

$IP_\alpha = [m - h ; m + h]$.

où h : précision vu au IC

$P(\bar{x} \in [m - h ; m + h]) = 1 - \alpha \simeq 95\%$

i/ IP:

si $\bar{x} \in IP \Rightarrow H_0$ non rejetée au risque α

si $\bar{x} \notin IP \Rightarrow H_0$ est rejetée.

ii/ IC:

si $m \notin IC_\alpha(m) \Rightarrow H_0$ est rejetée au risque α

si $m \in IC \Rightarrow H_0$ non rejetée au risque α

2ème méthode:

Basée sur l'écart-type.

a/ σ connu :

$$Z = \frac{|m - \bar{x}|}{\sigma / \sqrt{n}}$$

si $n > 30$ $Z \rightsquigarrow N(0,1)$.

si $n \leq 30$ $Z \rightsquigarrow N(0,1)$.

$X \rightsquigarrow N(m, \sigma^2)$

b/ σ inconnue :

$$Z = \frac{|m - \bar{X}|}{S/\sqrt{n-1}}$$

si $n > 30$ $\quad Z \rightsquigarrow N(0,1)$.

si $n \leq 30$ $\quad Z \rightsquigarrow St_{n-1}$

$X \rightsquigarrow N(m, \sigma^2)$

Quand est-ce qu'on rejette ou non H_0 ?

$H_0 : m = \bar{x} \Leftrightarrow d = |m - \bar{x}| \simeq 0$

différence est négligeable $\leftarrow$ | $\rightarrow$ différence est grande

$\Rightarrow H_0$ non rejetée.

$Z \leq q_{1-\frac{\alpha}{2}}$

$Z \leq St_{n-1}(\alpha)$

$\Rightarrow H_0$ rejetée.

$Z > q_{1-\frac{\alpha}{2}}$

$Z > St_{n-1}(\alpha)$

$q_{1-\frac{\alpha}{2}}$

$St_{n-1}(\alpha)$

Conclusion :

σ inconnue :

- $n > 30 \rightarrow$ si $Z_{obs} > q_{1-\frac{\alpha}{2}} \Rightarrow H_0$ rejetée au risque α
- $n \leq 30 \rightarrow$ si $Z_{obs} > t_{n-1}(\alpha) \Rightarrow H_0$ rejetée au risque α

σ connue :

- $n > 30$
- $n \leq 30$, $X \rightsquigarrow N(m, \sigma^2)$

$\rightarrow$ si $Z_{obs} > q_{1-\frac{\alpha}{2}} \Rightarrow H_0$ rejetée au risque α

si $Z_{obs} \leq q_{1-\frac{\alpha}{2}} \Rightarrow H_0$ non rejetée au risque α

II/ Test pour la proportion :

Il s'agit de comparer une proportion théorique connue p_0 à une proportion observée p.

$H_0 : p = p_0$

$H_1 : p \neq p_0$

H_0 : l'éch. est représentatif de la pop.

1ère méthode : $n > 30$; $np_0 \geq 5$ et $nq_0 \geq 5$

i/ $IP = [p_0 - h ; p_0 + h]$.

$$h = q_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0 q_0}{n}}$$

si $p \notin IP \Rightarrow H_0$ rejetée au risque α.

ii/ $IC = [p - h ; p + h]$

$$h = q_{1-\frac{\alpha}{2}} \sqrt{\frac{pq}{n}}$$

si $p_0 \notin IC \Rightarrow H_0$ rejetée au risque α.

2ème méthode : l'écart-type.

$$Z = \frac{|p - p_0|}{\sqrt{\frac{p_0 q_0}{n}}} \rightsquigarrow N(0,1) \quad n > 30 \; np_0 \geq 5 \text{ et } nq_0 \geq 5$$

si $Z_{obs} > q_{1-\frac{\alpha}{2}} \Rightarrow H_0$ est rejetée au risque α

N.B. :

Rejeter H_0 au risque $\alpha \Rightarrow$ rejeter H_0 au risque α' pour $\forall \alpha' > \alpha$.

III/ Test du χ^2:

Il s'agit de comparer une distribution observée à une distribution théorique: c'est la généralisation du cas précédent

	x_i	x_1	x_2	...	x_k	Σ
théo.	P_{0i}	P_{01}	P_{02}	...	P_{0k}	1
obs.	P_i	P_1	P_2	...	P_k	1

$p_i = P(X = x_i)$

k = nbre de classes (modalités)

H_0: la distribution théorique (pop) est conforme à la distribution observée (éch).

H_0: $p_{0i} = p_i \quad \forall i = 1, 2, \dots, k.$

$$\chi^2_{cal} = \sum_{i=1}^{k} \frac{(\theta_i - c_i)^2}{c_i} \quad \forall c_i \geq 5.$$

avec: θ_i = effectif observé de la classe i

c_i = effectif théorique (calculé) de la classe i.

$c_i = n p_{0i}$ — n = taille de l'éch.; p_{0i} = la proportion théorique de la classe i.

Classes	θ_i	p_{0i}	$c_i = np_{0i}$	$\frac{(\theta_i - c_i)^2}{c_i}$
⋮	⋮	⋮	⋮	⋮
Σ	n	1	n	χ^2_{cal}

$\chi^2_{cal} \rightsquigarrow \chi^2_\nu$ où $\nu = k - 1$

si $\chi^2_{cal} > \chi^2_{k-1}(\alpha) \Rightarrow H_0$ rejetée (α).

où $\chi^2_{k-1}(\alpha)$ lu sur la table du χ^2 correspondant à α et $\nu = ddl = k - 1$

Classes	θ_i	p_{0i}	$c_i = np_{0i}$	$\frac{(\theta_i - c_i)^2}{c_i}$
O	35	0,45	45	
A	35	0,35	35	
B	20	0,16	16	$\frac{(30-20)^2}{20}$
AB	10 (30)	0,04	4 (20)	
Σ	n 100	1	$n = 100$	$\chi^2_{cal} = 7,22$

N.B.:

$k > 2$ et $\exists c_i < 5$

Dans ce cas, on fait un regroupement de classes.

Les classes à regrouper doivent être successives.

On regroupera autant de classes que nécessaire pour que tous les $c_i \geq 5$.

Soit k' le nbre de classes après regroupement

$\nu = ddl = k' - 1$

$k = 4 \rightarrow k' = 3 \Rightarrow \nu = k' - 1 = 2$

si $\chi^2_{cal} > \chi^2_2(0,05) = 5,95$.

$\Rightarrow H_0$ rejetée au risque 5%.

$k = 2$ et $\exists c_i < 5$.

Pas de rejet

On procède à la «correction de Yates»:

$$\chi^2_{cal} = \sum_{i=1}^{2} \frac{(|\theta_i - c_i| - \frac{1}{2})^2}{c_i}$$

si $\chi^2_{cal} > \chi^2_1(\alpha) \Rightarrow H_0$ rejetée (α)

Cas particulier de 2 classes:

H_0: $p_{o_1} = p_1$ et $p_{o_2} = p_2$

$1 - p_{o_1} = 1 - p_1$ $\quad 1 - p_{o_2} = 1 - p_2$

$p_{o_2} = p_2$

x_i	x_1	x_2
p_{o_i}	p_{o_1}	p_{o_2}
p_i	p_1	p_2

Dans ce cas, le test de conformité du χ^2 est équivalent au test de conformité de l'écart réduit pour la proportion:

Exo:

On lance une pièce de monnaie fois, on obtient 45 pile?

- La pièce est-elle équilibrée?

H_0: la pièce est équilibrée.

45 pile $\rightarrow p = \frac{45}{100} = 0{,}45$ $\quad p_o = \frac{1}{2}$

H_0: $p = p_o$

1ère méthode: l'écart type: $\varepsilon = 1$

$$\varepsilon = \frac{|p - p_o|}{\sqrt{\frac{p_o q_o}{n}}} \rightsquigarrow N(0,1) \quad np_o = 50 > 5 \text{ et } nq_o = 50 > 5$$

$$\varepsilon = \frac{|0{,}45 - 0{,}5|}{\sqrt{\frac{0{,}5 \cdot 0{,}5}{100}}} = 1 \qquad \alpha = 5\%$$

$q_{1-\frac{\alpha}{2}} = 1{,}96$

$1 < 1{,}96 \Rightarrow H_0$ non rejetée à 5%

2e méthode: test du χ^2:

H_0:

$$\chi^2_{cal} = \sum_{i=1}^{2} \frac{(\theta_i - c_i)^2}{c_i}$$

Classes	θ_i	p_{o_i}	c_i	$\frac{(\theta_i - c_i)^2}{c_i}$
Pile	45	$\frac{1}{2}$	50	$\frac{1}{2}$
Face	55	$\frac{1}{2}$	50	$\frac{1}{2}$
Σ	100	1	100	$\chi^2_{cal} = 1$

$k = 2 \rightarrow \nu = 1$

$\chi^2_1(0{,}05) = 3{,}84$

$\chi^2_{cal} = 1 < 3{,}84 \Rightarrow H_0$ non rejetée au risque 5%

N.B.: $\varepsilon^2 = \chi^2_{cal}$

IV/ Test d'ajustement du χ^2:

$p_{o_i} = P(X = i)$ $\quad X \rightsquigarrow$ loi usuelle.

ex: $X \rightsquigarrow B(n,p)$.

$p_{o_i} = P(X = i) = C_n^i \, p^i (1-p)^{n-i} \quad \forall i = 0,1,\dots n$

Il s'agit toujours de comparer une distribution observée à une distribution théorique, mais dans ce test, la variable d'interêt X est de loi usuelle. Alors, on calculera $p_{o_i} = P(X = i)$ selon la loi théorique de la pop.

ex: (voir plus haut).

H_0: La loi observée ajuste la loi théorique.

$$\chi^2_{cal} = \sum_{i=1}^{k} \frac{(\theta_i - c_i)^2}{c_i} \qquad \forall c_i \geq 5.$$

Si: $\chi^2_{cal} > \chi^2_{\nu}(\alpha) \Rightarrow H_0$ rejetée au risque (α)

$\nu = ddl = k' - 1 - r$

k': nbre de classes après regroupement

r: nbre de paramètres inconnus de la loi théorique.

ex: $X \rightsquigarrow B(n,p)$ n connu

$$r = \begin{cases} 0 \text{ si } p \text{ connu} \\ 1 \text{ si } p \text{ inconnu} \end{cases}$$

p inconnu, alors à estimer. $\hat{m} = \bar{x}$

$m = E(X) = np$. $\hat{\sigma}^2 = \frac{n}{n-1} s^2$

$\hat{m} = \bar{x}$

Donc: $\hat{m} = n\hat{p} = \bar{x}$

$\Rightarrow \hat{p} = \frac{\bar{x}}{n}$

N.B.: $\bar{x} = \frac{1}{N} \sum_{i=1}^{k} \theta_i x_i$ avec n_i : N: taille de l'éch.

classes	θ_i	p_{0i} ou $\hat{p}_{0i}$	$c_i = N p_{0i}$	$\frac{(\theta_i - c_i)^2}{c_i}$
$i=0$				
$i=1$				
$\vdots$				
$i=n$				
$\sum$	N	1	N	χ^2_{cal}

$\hat{p}_{0i} = P(X=i) = C_n^i \hat{p}^i (1-\hat{p})^{n-i}$

ex2: $X \rightsquigarrow \mathcal{P}(\lambda) \Leftrightarrow P(X=i) = e^{-\lambda} \frac{\lambda^i}{i!}$ $(\forall i \in \mathbb{N})$.

$$r = \begin{cases} 0 \text{ si } \lambda \text{ connu} \\ 1 \text{ si } \lambda \text{ inconnu} \end{cases}$$

λ inconnu, alors à estimer.

$m = E(X) = \lambda \Rightarrow \hat{\lambda} = \bar{x}$

$\hat{m} = \hat{\lambda}$

classes	θ_i	p_{0i} ou $\hat{p}_{0i}$
$i=0$	θ_0	p_{a_0}
$i=1$	θ_1	p_{0_2}
$\vdots$	$\vdots$	$\vdots$
$i=k$	θ_k	p_{a_k}
$i>k$	0	$1 - \sum_{i=0}^{k} p_{0i}$
	N	1

Pour calculer p_{0i}, on utilise la relation de récurrence.

$P(X=0) = e^{-\lambda}$

$P(X=i+1) = \frac{\lambda}{i+1} \cdot P(X=i)$.

ex3:

$X \rightsquigarrow N(m, \sigma^2)$ $\mathcal{X} = \mathbb{R}$.

$$r = \begin{cases} 0 \text{ si } m \text{ et } \sigma^2 \text{ connus} \\ 1 \text{ si } m \text{ ou } \sigma^2 \text{ connu} \\ 2 \text{ si } m \text{ et } \sigma^2 \text{ inconnus} \end{cases}$$

$\hat{m} = \bar{x}$ $\hat{\sigma}^2 = \frac{n}{n-1} s^2$

classes	θ_i	p_{0i}
$]-\infty; a_0[$	0	$P(X<a_0)$
$[a_0; a_1[$	θ_1	$P(a_0 \le X < a_1)$
$[a_1; a_2[$	θ_2	$\vdots$
$\vdots$	$\vdots$	$\vdots$
$[a_{k-1}; a_k[$	θ_k	$\vdots$
$[a_k; +\infty[$	0	$P(X \ge a_k)$
	N	1

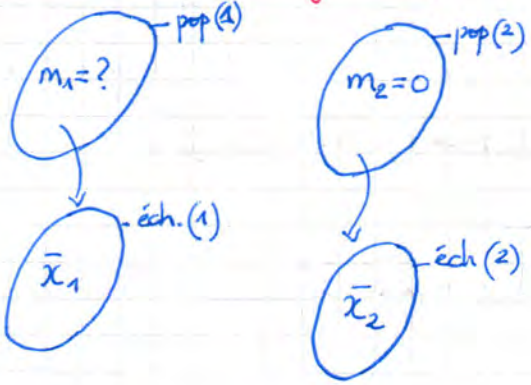
ex4: $X \rightsquigarrow \mathcal{E}(\lambda) \Leftrightarrow f(x) = \lambda e^{-\lambda x}$ si $\lambda > 0$

$m = E(X) = \frac{1}{\lambda}$

$\Rightarrow \hat{\lambda} = \frac{1}{\bar{x}}$

$\hat{m} = \bar{x}$

C/ Tests d'homogénéité:



Sur: www.la-faculte.net Espace E-learning pour apprentissage gratuit online Pour utilisation Non-lucrative

Il s'agit de comparer 2 pop. à travers leurs paramètres, mais les paramètres des pop. ne sont pas connus. Alors, on extrait un éch. représentatif de chaque pop. et on comparera les paramètres observés de ces éch.

I/ Comparaison de 2 moyennes:

$H_0 : m_1 = m_2$ vs. $H_1 : m_1 \neq m_2$

H_0 : il n'y a pas de différence significative entre les 2 moyennes.

H_0 : les éch. proviennent de la même pop. ou de 2 pop. identiques.

1ère méthode: recouvrement des IC.

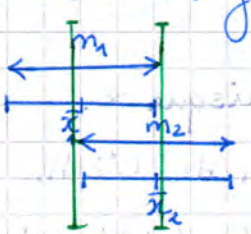
$I_1 = IC_\alpha(m_1) = [\bar{x}_1 - h_1 ; \bar{x}_1 + h_1]$

$I_2 = IC_\alpha(m_2) = [\bar{x}_2 - h_2 ; \bar{x}_2 + h_2]$

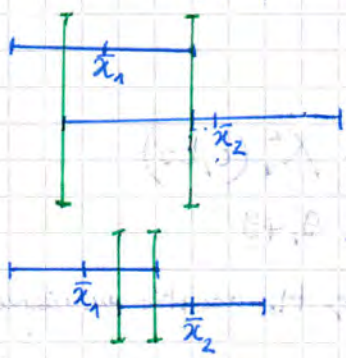
• $I_1 \cap I_2$:

•• $I_1 \cap I_2$ avec $\bar{x}_1$ et $\bar{x}_2 \in I_1 \cap I_2$

$\Rightarrow H_0$ non rejetée au risque (α).



••• $I_1 \cap I_2 \neq \emptyset$ mais $\bar{x}_1$ ou $\bar{x}_2 \notin I_1 \cap I_2 \Rightarrow ?$



On ne peut rien dire, il faut faire le test de l'écart réduit.

2ème méthode: test de l'écart réduit

H_0 :

Calcul:

Si $n_1 > 30$ et $n_2 > 30$:

On supposera σ_1^2 et σ_2^2 inconnues.

$$\varepsilon = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$\bar{X}_1 \rightsquigarrow N\left(m_1 ; \frac{\sigma_1^2}{n_1}\right)$

$\bar{X}_2 \rightsquigarrow N\left(m_2 ; \frac{\sigma_1^2}{n_2}\right)$

$\bar{X}_1 - \bar{X}_2 \rightsquigarrow N\left(m_1 - m_2 ; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

Si $\varepsilon > q_{1-\frac{\alpha}{2}} \Rightarrow H_0$ rejetée au risque α

Si $n_1 \leq 30$ ou $n_2 \leq 30$:

On doit supposer que $X_1 \rightsquigarrow N(m_1, \sigma_1^2)$ et $X_2 \rightsquigarrow N(m_2, \sigma_2^2)$.

$\sigma_1^2 = \sigma_2^2 = \sigma^2$ inconnu

Alors : $\varepsilon = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$ où $S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$

$\hat{\sigma}^2 = S^2$

$\varepsilon \rightsquigarrow St_\nu$

$n_{éch.}$ $n_1 - 1$

2e éch. $\nu_2 = n_2 - 1$

$n_1 + n_2 - 2$

$\nu = n_1 + n_2 - 2$

facadm16@gmail.com Participez à "Q&R rapide" pour mieux préparer vos examens 2017/2018

Sur: www.la-faculte.net Espace E-learning pour apprentissage gratuit online Pour utilisation Non-lucrative

suppose la normalité.

$$\varepsilon = \frac{\sqrt{n-2}\,|r_{xy}|}{\sqrt{1-r_{xy}^2}}$$

Sous H_0 $\varepsilon \rightsquigarrow St_{n-2}$.

Si $\varepsilon > t_{n-2}(\alpha) \Rightarrow H_0$ rejetée (α).

Exercices :

Exo 5 :

$n = 100$

$\bar{x} = 20$; $\bar{y} = 12$.

$\bar{x^2} = \frac{1}{n}\sum x_i^2$; $\bar{y} = \frac{1}{n}\sum y_i^2 = 153$.

$= 436$

$\bar{xy} = \frac{1}{n}\sum x_i y_i = 254{,}4$.

1/ $r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$ $\quad S_{xy} = \frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}$.

$= 254{,}4 - 20.12$.

$r_{xy} = \frac{14{,}4}{6.3} = 0{,}8$. $\quad S_{xy} = 14{,}4$.

$S_x^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2$

$= 436 - 20^2$

$S_x^2 = 36$

$S_y^2 = \frac{1}{n}\sum y_i^2 - \bar{y}^2$

$= 153 - 12^2$

$S_y^2 = 9$

2/ $Dy(x) \rightarrow \hat{y} = \hat{a}x + \hat{b}$

$$\begin{cases} \hat{a} = \frac{S_{xy}}{S_x^2} = \frac{14{,}4}{36} \\ \hat{b} = \bar{y} - \hat{a}\bar{x}. \end{cases}$$

$= 12 - \frac{14{,}4}{36} \cdot 20 = .$

$Dx(y) \rightarrow \hat{x} = \hat{\alpha}y + \hat{\beta}$.

$\hat{\alpha} = \frac{S_{xy}}{S_y^2} = \frac{14{,}4}{9} =$

$\hat{\beta} = \bar{x} - \hat{\alpha}\bar{y} = 20 - \frac{14{,}4}{9} \times 12 = \ldots$

Exo 6 :

$\bar{x} = 25$ $\quad S_x^2 = 25$.

$\bar{y} = 3100$ $\quad S_y^2 = 10^4$.

$S_{xy} = 450$.

1/ $r_{xy} = \frac{S_{xy}}{S_y \cdot S_x} = \frac{450}{5.100} = 0{,}9$.

2/ $r_{xy} = 0{,}9 \rightarrow r_{xy}^2 = 0{,}81 > \frac{3}{4} \Rightarrow$ Oui.

3/ $r_{xy} = 0{,}9 > 0 \Rightarrow$ X et Y croient dans le même sens.

$x_1 = 20 < x_2 = 30 \Rightarrow y_1 < y_2$.

Exo 7 :

1/ $r_{xy} = 0{,}94$.

2/ Oui (voir exo 6).

3/ $Dy(x) \rightarrow \hat{y} = \hat{a}x + \hat{b}$.

$\hat{y} = 0{,}33x + 1{,}32$.

4/ Oui car $r_{xy} = 0{,}94 \approx 1$.

et $x_0 = 13 \rightarrow y_{13} = 0{,}33 \times 13 + 1{,}32$.

Exo 8 :

éch. 1 $n_1 = 400$, $r_1 = 0{,}2$

éch. 2 $n_2 = 625$, $r_2 = 0{,}38$.

1/ éch. 1 $r_1 = 0{,}2 \neq 1$ et $r_1 = 0{,}2 \neq 0$.

Test de conformité.

H_0 : A et B ne sont pas corélés.

$\varepsilon = \frac{\sqrt{n-2}\,|r_0|}{\sqrt{\quad}}$ $\quad \sqrt{\quad}$

facadm16@gmail.com Participez à "Q&R rapide" pour mieux préparer vos examens 2017/2018

Sur: www.la-faculte.net Espace E-learning pour apprentissage gratuit online Pour utilisation Non-lucrative

$C_{12} = \frac{70 \times 50}{100} = 35$.

Dans un tableau 2-2, il suffit de calculer un seul des C_{ij}, et les autres se déduisent par addition.

Test.

H_0 : les 2 caractères sont indépendants.
⇒ le tabagisme n'influe pas sur le cancer du poumon.

ex:

$$\chi^2_{cal} = \sum_i \sum_j \frac{(o_{ij} - c_{ij})^2}{c_{ij}} \quad \forall c_{ij} \geq 5$$

$$\chi^2_{cal} \rightsquigarrow \chi^2_\gamma$$

$$\gamma = (r-1) \times (l-1).$$

Si $\chi^2_{cal} > \chi^2_\gamma(\alpha)$

⇒ H_0 rejetée au risque α.

Cas où $c_{ij} < 5$:

Si r = nbre de classes. $r > 2$ et $\exists c_{ij} < 5$

⇒ regroupement de classes.

⇒ $\gamma = (r' - 1) \times (l - 1)$.

Ech \ Classes	1	2	3	4
A			2,3	3,2
B				
C				
D				

$r = 2$ et $l = 2$ et $\exists c_{ij} < 5$

⇒ Yates $\chi^2_{cal} = \sum_i \sum_j \frac{(|o_{ij} - c_{ij}| - \frac{1}{2})^2}{c_{ij}}$

Cas particulier d'un tableau 2x2:

ex précédent.

H_0 : le tabac n'influe pas sur le cancer du poumon:

$P(C/F) = P(C/\bar{F})$ ou $P(\bar{C}/F) = P(\bar{C}/\bar{F})$

H_0 : $p_1 = p_2$ ⇒ $1 - p_1 = 1 - p_2$

$q_1 = q_2$

Dans ce cas, le test d'ind. du χ^2 est équivalent au test de l'écart réduit pour la comparaison de 2 proportions données

$$\varepsilon^2 = \chi^2_{cal}$$

Test de l'écart réduit:

$n_1 p_1 = 40 > 5$ $\quad n_1 q_1 = 30 > 5$

$n_2 p_2 = 10 > 5$ $\quad n_2 q_2 = 20 > 30$

$$p_1 = P(C/F) = \frac{|C \cap F|}{|F|} = \frac{40}{70}$$

$$p_2 = P(C/\bar{F}) = \frac{|C \cap \bar{F}|}{|\bar{F}|} = \frac{10}{30}.$$

$$\varepsilon = \frac{|p_1 - p_2|}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad p = \frac{40+10}{100} = \frac{1}{2}$$

$$\varepsilon = \frac{|4/7 - 1/3|}{\sqrt{\frac{1}{2} \cdot \frac{1}{2}\left(\frac{1}{70} + \frac{1}{30}\right)}}$$

Exercices:

Exo 3:

IG \ MT	R	L	Σ
A	2 / 8	38 / 32	
N	78 / 72	282 / 288	
Σ	80	320	400

H_0 : il n'y a pas de liaison (il y a indép.) entre la MT et IG.

$p_1 = P(N/L)$ $\quad p_2 = P(N/R)$.

ou $1 - p_1 = P(A/L)$ $\quad 1 - p_2 = P(A/R)$.

H_0 : $p_1 = p_2$ ⇔ H_0 : $q_1 = q_2$

$$q_1 = P(A/L) = \frac{|A \cap L|}{|L|} = \frac{38}{320}$$

$$q_2 = P(A/R) = \frac{|A \cap R|}{|R|} = \frac{2}{80}$$

ou $p_1 = P(N/L) = \frac{|N \cap L|}{|L|} = \frac{282}{320}$

$p_2 = P(N/R) = \frac{|N \cap R|}{|R|} = \frac{78}{80}$

facadm16@gmail.com Participez à "Q&R rapide" pour mieux préparer vos examens 2017/2018

$H_0 : p_1 = p_2$ → IC si $n_1 > 30$ et $n_2 > 30$
→ Écart réduit $n_1 p_1 \geq 5$ $n_1 q_1 \geq 5$
$n_2 p_2 \geq 5$ $n_2 q_2 \geq 5$
→ χ^2_{cal} pas de conditions

On a $n_1 p_1 = k_1 = 2 < 5 \Rightarrow$ pas d'écart réduit

$\chi^2_{cal} = 6,25$.

$\gamma = 1 \Rightarrow \chi^2_{cal}(0,05) = 3,84$.

H_0 rejetée à 5%

E/ Tests de corrélation:

Il s'agit d'étudier l'existence d'une liaison entre 2 caractères quantitatifs

ex: X: poids d'une personne.
Y: Taille de cette personne.

On dispose d'un éch. de taille n. Cet éch. est représenté par un nuage de points

C'est la forme du nuage de points qui déterminera le type de liaison entre X et Y.

On s'intéressera à une relation linéaire entre X et Y. $Y = aX + b$.

Coefficient de corrélation:

Déf 1: le coefficient de corrélation (linéaire) donne le degré de liaison entre X et Y, noté r_{xy} ou ρ_{xy}.

Déf 2:

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}.$$

où S_x et S_y sont respectivement les écart-types de X et Y, et S_{xy} est la covariance de (X, Y).

$$S_{xy} = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})$$

N.B.:

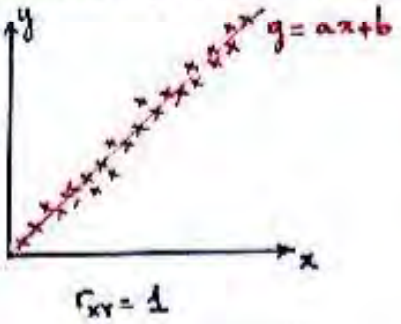
$$S_{xx} = \frac{1}{n} (x_i - \bar{x})^2 = S_x^2 = \frac{1}{n} \sum x_i y_i - \bar{x}\bar{y}.$$
$$= \frac{1}{n} \sum x_i^2 - \bar{x}^2$$

Propriétés:

- $|r_{xy}| \leq 1$
- Si $|r_{xy}| \simeq 1 \Rightarrow$ très forte liaison (linéaire) entre X et Y.

$Y = aX + b$.

X et Y corrélés $\Rightarrow$ X et Y non indépendants.



$r_{xy} = 1$

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}.$$

$|r_{xy}| \simeq 1$ si et seulement si $r_{xy}^2 \geq \frac{3}{4}$

- Si $r_{xy} > 0$ ($\Leftrightarrow S_{xy} > 0$) $\Rightarrow$ X et Y croient dans le même sens.

Si $X \nearrow = Y \nearrow$ et $X \searrow = Y \searrow$.

Si $r_{xy} < 0$ ($\Leftrightarrow S_{xy} < 0$).

$\Rightarrow$ X et Y croient au sens différents.

Sur: www.la-faculte.net Espace E-learning pour apprentissage gratuit online Pour utilisation Non-lucrative

• Si $|r_{xy}| \simeq 0$ => Il y a une très faible corrélation entre X et Y

=> X et Y non corrélés => X et Y indépendants

Si r_{xy} n'est ni assez proche de 1 ni de 0, alors, il faut faire test statistique.

Droites de régression:

$|r_{xy}| \simeq 1 \rightarrow Y = ax+b$

On spécifiera la liaison, c'est à dire: on calculera a et b.

Il existe 2 droites de régression:

- La première droite de régression de Y en X, notée: $\hat{y} = \hat{a}x + \hat{b}$.

C'est la droite pour laquelle la somme des écarts des points à la droite parallèle à (Oy) est minimum.

On obtient:

$$\hat{a} = \frac{S_{xy}}{S_x^2} \quad \text{et} \quad \hat{b} = \bar{y} - \hat{a}\bar{x}.$$

- La droite de régression de X en Y.

$D_x(y) \rightarrow \hat{x} = \hat{\alpha}y + \hat{\beta}$

C'est la droite pour laquelle la somme des écarts des points à la droite parallèle à (Ox) est minimum.

On obtient:

$$\hat{\alpha} = \frac{S_{xy}}{S_y^2} \quad \text{et} \quad \hat{\beta} = \bar{x} - \hat{\alpha}\bar{y}.$$

Les 2 droites se coupent en un point $G(\bar{x}, \bar{y})$ appelé centre de gravité.

La droite de régression de Y en X permet de prévoir une valeur de Y pour une valeur de x donnée

$D_{y(x)} \rightarrow \hat{y} = \hat{a}x + \hat{b}$.

x_0 donné $\rightarrow y_{x_0} = \hat{a}x_0 + \hat{b}$.

$D_x(y) \rightarrow \hat{x} = \hat{\alpha}y + \hat{\beta}$.

Elle de prédire une valeur de x pour une valeur de y donnée.

Test de corrélation:

Si $|r_{xy}|$ n'est ni proche de 1 ni proche de 0, on procède à un test de conformité.

H_0: X et Y indépendants.

H_1: X et Y non corrélés.

H_0: $r_{xy} = 0$ (↓ éch. ; ↓ thé.)

$$\varepsilon = \frac{\sqrt{n-2}\,|r_{xy}|}{\sqrt{1-r_{xy}^2}}$$

Sous H_0 $\varepsilon \rightsquigarrow St_{n-2}$.

Si $\varepsilon > t_{n-2}(\alpha)$ => H_0 rejetée (α).

Test d'homogénéité:

Il permet de comparer 2 coefficients de corrélation observés.

éch. 1, taille n_1, coefficient de corrélation r_1.

éch. 2, taille n_2, coefficient de corrélation r_2.

H_0: les 2 éch. proviennent de la même pop.

$$\varepsilon = \frac{|r_1 - r_2|}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

Sous H_0: $\varepsilon \rightsquigarrow N(0,1)$.

Rejeter H_0 si $\varepsilon > q_{1-\frac{\alpha}{2}}$

facadm16@gmail.com Participez à "Q&R rapide" pour mieux préparer vos examens 2017/2018